

Arch Sex Behav (2012) 41:13–29
DOI 10.1007/s10508-010-9675-3

ORIGINAL PAPER

Sexual Attraction to Others: A Comparison of Two Models of Alloerotic Responding in Men

Ray Blanchard · Michael E. Kuban · Thomas Blak · Philip E. Klassen · Robert Dickey · James M. Cantor

Received: 9 February 2010 / Revised: 12 May 2010 / Accepted: 11 August 2010 / Published online: 17 September 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract The penile response profiles of homosexual and heterosexual pedophiles, hebephiles, and teleiophiles to laboratory stimuli depicting male and female children and adults may be conceptualized as a series of overlapping stimulus generalization gradients. This study used such profile data to compare two models of alloerotic responding (sexual responding to other people) in men. The first model was based on the notion that men respond to a potential sexual object as a compound stimulus made up of an age component and a gender component. The second model was based on the notion that men respond to a potential sexual object as a gestalt, which they evaluate in terms of global similarity to other potential sexual objects. The analytic strategy was to compare the accuracy of these models in predicting a man's penile response to each of his less arousing (nonpreferred) stimulus categories from his response to his most arousing (preferred) stimulus category. Both models based their predictions on the degree of dissimilarity between the preferred stimulus category and a given nonpreferred stimulus category, but each model used its own measure of dissimilarity. According to the first model ("summation model"), penile response should vary inversely as the sum of stimulus differences on separate dimensions of age and gender. According to the second model ("bipolar model"), penile response should vary inversely as the distance between stimulus categories on a single, bipolar dimension of morphological similarity—a dimension on which children are

located near the middle, and adult men and women are located at opposite ends. The subjects were 2,278 male patients referred to a specialty clinic for phallometric assessment of their erotic preferences. Comparisons of goodness of fit to the observed data favored the unidimensional bipolar model.

Keywords Homosexuality · Heterosexuality · Bisexuality · Pedophilia · Hebephilia · Teleiophilia · Penile plethysmography · Mathematical models · Stimulus generalization gradient

Introduction

The term *alloerotic* means the opposite of *autoerotic*, that is, it denotes sexual attraction and sexual response to other people. There are two well established facts about alloerotic responding in men. The first is that the great majority of men respond most strongly to persons of a particular age (or age-range) and gender. The second is that men also respond sexually to persons outside their preferred category, in some rough proportion to their similarity to persons inside the preferred category. What is not known is exactly how the factors of age and gender combine to determine the relative attractiveness of people in nonpreferred categories. That is the topic of the present research.

One possible model of alloerotic responding is to conceptualize age and gender as separate stimulus dimensions, and to hypothesize that differences in gender and differences in age between men's preferred and nonpreferred stimulus categories independently diminish their responses to persons in nonpreferred categories. On this view, homosexual pedophiles, for example, are less attracted to prepubescent females than to prepubescent males because prepubescent females have the

A version of this article was presented at the University of Lethbridge Workshop, *What Is Sexual Orientation?*, Lethbridge, Alberta, Canada, June 2010.

R. Blanchard (✉) · M. E. Kuban · T. Blak · P. E. Klassen · R. Dickey · J. M. Cantor
Law and Mental Health Program, Centre for Addiction and Mental Health, 250 College Street, Toronto, ON M5T 1R8, Canada
e-mail: ray_blanchard@camh.net

right age but the wrong gender, and they are less attracted to adult males because they have the right gender but the wrong age. They are least attracted to adult females because they have the wrong age and the wrong gender. We will refer to this possibility as the summation model. The term *summation* refers to the notion that sexual response to any nonpreferred stimulus category varies inversely as the sum of that category's differences from the preferred category.

There is nothing obviously false in the above model, but there is one fact that suggests one consider alternatives. In the context of human development, age and gender are not really separate dimensions. Sexual dimorphism is partly a function of age. At birth, human males and females look very little different, except for the appearance of the external genitalia. The difference in outward appearance increases with age, accelerating at puberty, and reaching its maximum sometime in adulthood. This suggests the possibility of an alternative model: Men respond sexually as if they perceive other humans as points along a single, bipolar dimension of morphological similarity—a dimension in which children are located near the middle, and adult men and women are located at opposite ends. On this view, men's sexual response to persons in a nonpreferred stimulus category is a simple function of the “distance” between that category and their preferred category.

The first goal of the present studies was to investigate which conceptualization—the summation model or the bipolar dimension model—best describes the sexual responding of men in general, where “men in general” are represented by a heterogeneous sample of adult males with a wide range of erotic preferences. The second goal was to investigate whether the same model works best for each of six subgroups of men: homosexual pedophiles (most attracted to prepubescent boys), heterosexual pedophiles (most attracted to prepubescent girls), homosexual hebephiles (most attracted to pubescent boys), heterosexual hebephiles (most attracted to pubescent girls), homosexual teleiophiles (most attracted to physically mature men), and heterosexual teleiophiles (most attracted to physically mature women).

The method used for quantifying sexual response in these studies was phallometric testing. This is a psychophysiological technique for assessing erotic interests in male adults and adolescents. In phallometric tests for gender and age orientation, the individual's penile blood volume is monitored while he is presented with a standardized set of laboratory stimuli depicting male and female children, pubescents, and adults. Increases in the patient's penile blood volume (i.e., degrees of penile erection) are used as the measure of his attraction to different classes of persons.¹

The present studies are, in some ways, an extension of phallometric research initiated by Freund, Langevin, Cibiri, and Zajac (1973). These authors plotted the penile responses of heterosexual and homosexual teleiophiles to stimulus categories representing persons of both genders and varying ages. They arranged the stimulus categories along the X-axis in the following order: adult females, pubescent females, older prepubescent (8- to 11-year-old) females, younger prepubescent (6- to 8-year-old) females, younger prepubescent (6- to 8-year-old) males, older prepubescent (9- to 11-year-old) males, pubescent males, and adult males (see Freund et al., 1973, Fig. 1). In other words, the youngest children were located near the middle of the axis, and adult males and females were located at opposite ends. It is unknowable, at this time, whether they had something like the single, bipolar dimension model in mind or whether (more likely) they simply wanted to illustrate a point about the comparability of heterosexual and homosexual teleiophiles. In any event, the results showed that their (self-reported) heterosexual teleiophiles responded most to adult females, less to the other three categories of females—penile response decreasing as the age of females decreased—and little or not at all to the four categories of males. Strikingly symmetrical results were found for the homosexual teleiophiles, whose response profile showed its maximum value for adult males.

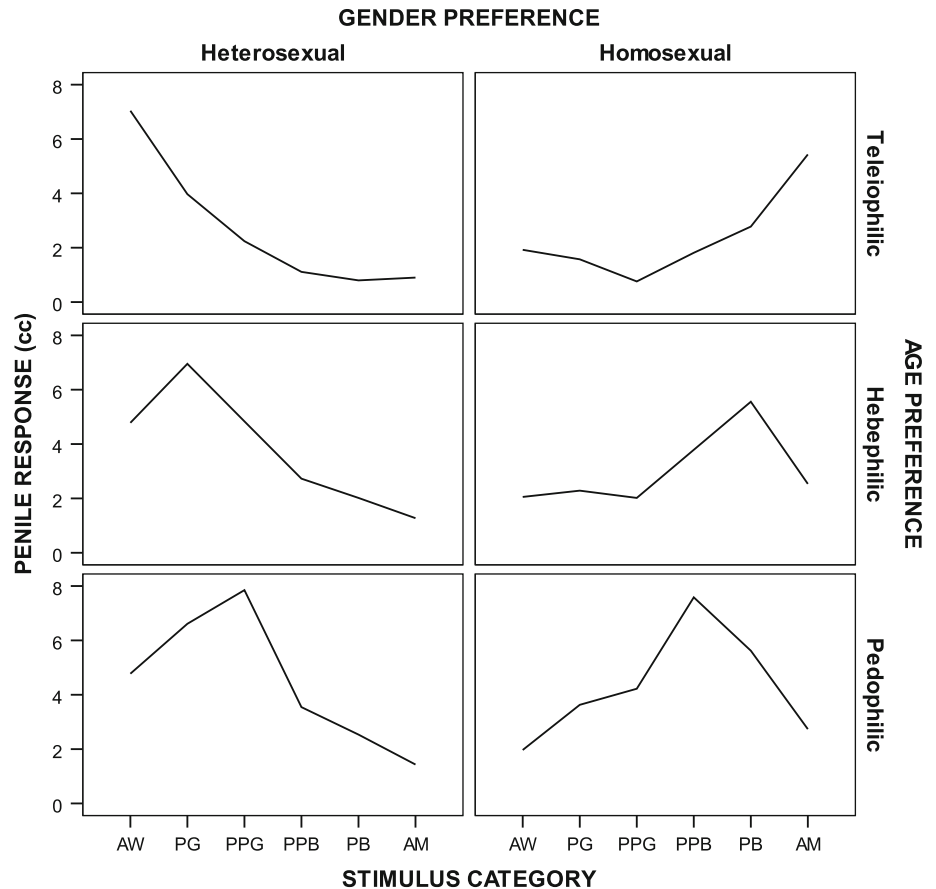
The phallometric response profiles published by Freund et al. (1973) and by other investigators (e.g., Blanchard et al., 2009a, b; Frenzel & Lang, 1989; Freund, McKnight, Langevin, & Cibiri, 1972; Lykins et al., 2010a) are strongly reminiscent of the stimulus generalization gradients studied by experimental psychologists. A stimulus generalization gradient is a graphic depiction of the extent to which behavior that is most strongly elicited by a given stimulus is also elicited by stimuli that are similar but not identical to it. In experimental psychology, the maximally excitatory stimulus is usually established with classical or operant conditioning. For example, a pigeon trained to peck a key for food when it is illuminated with light of a particular wavelength will peck the key at lower rates when it is illuminated with other wavelengths; the rate of pecking is directly related to the proximity of the testing wavelength to the training wavelength (e.g., Blough, 1969). We do not mean to imply, by this comparison, that erotic preferences are established by classical or operant conditioning. We mean, rather, to point out that phallometric response profiles are analogous to stimulus generalization gradients in that both are products of organisms' behavior in relation to perceived stimulus similarity.

Footnote 1 continued

notable breakdown of linearity: Very excitable (usually younger) males may respond with full erection to all age categories of females, even though their behavior outside of the laboratory and their self-report indicate a preferential attraction to one specific age-range. Such phallometric ceiling effects are quite rare, however, even in young subjects.

¹ For most cooperative and medically healthy individuals, the relation between penile tumescence and the hypothetical construct of sexual attraction may be treated as linear, or at least monotonic. There is one

Fig. 1 Phallometric response profiles of the six groups. Each group is shown in a separate panel. Abbreviations for stimulus categories: *AW* adult women, *PG* pubescent girls, *PPG* prepubescent girls, *PPB* prepubescent boys, *PB* pubescent boys, *AM* adult men



The available studies of phallometric profiles suggested that the orderliness of phallometric data—at least in the aggregate—would make them suitable for comparing the summation and bipolar dimension models of alloerotic responding in men. This orderliness also prompted the third goal of the present research, that is, to express the summation and bipolar dimension models in the form of competing equations intended to predict all the points on a man's phallometric response profile solely from the magnitude of his highest response—regardless of whether his highest response is to males or females, to children, pubescent, or adults.

Method

Subjects

Between November 1995 and October 2009, 3,166 male patients were administered the same phallometric test for erotic object (gender and age) preferences at the Kurt Freund Laboratory of the Centre for Addiction and Mental Health (Toronto, Ontario, Canada). The sources of the clinical referrals included parole and probation officers, prisons, defense lawyers, various institutions (ranging from group

homes for mentally retarded persons to regulatory bodies for health or educational professionals), and physicians in private practice. As would be expected from the preponderance of criminal justice sources, the majority of patients had one or more sexual offenses against children, adults, or both. Men who had no involvement with the criminal justice system and who initiated referrals through their physicians included patients who were unsure about their sexual orientation, patients concerned about hypersexuality or “sex addiction,” patients experiencing difficulties because of their excessive use of telephone sex lines or massage parlors, clinically obsessional patients with intrusive thoughts about unacceptable sexual behavior, and patients with paraphilic behaviors like masochism, fetishism, and transvestism. Subsets of these patients have been analyzed in previous studies, which report additional information about the patients' characteristics (e.g., Blanchard, Klassen, Dickey, Kuban, & Blak, 2001; Blanchard et al., 2007, 2009b).

The preliminary inclusion criteria for this study were that the patient had given informed consent for his assessment data to be used for research purposes, and that his sexual history data were complete and had been cross-checked at the time the data were retrieved for this study. The 2,725 subjects who satisfied the foregoing criteria were further reduced to a final

sample of 2,278 according to additional criteria described later. These 2,278 men had a mean age of 37.48 years ($SD = 13.21$) and a median education of high school graduation.

Materials and Measures

Sexual History

A standardized form, described in detail by Blanchard et al. (2009b), was used to record the patient's history of sexual offenses. Most of that information came from objective documents that accompanied his referral, for example, reports from probation and parole officers. The offense-history data were cross-checked against, and supplemented by, other information provided by the patient himself, including the number and nature of any additional sexual offenses that were admitted by the patient but for which he was never charged. The patient was also asked to rate his sexual attraction to persons in 12 gender–age categories (e.g., females aged 17 years or older, males aged 17 years or older, females aged 15–16 years, males aged 15–16 years, and so on) using a 5-point scale. The patient's information was solicited by the laboratory manager in a structured sexual history interview, which the manager conducted the same day he administered the phallometric test.

Phallometric Measurement

The Kurt Freund Laboratory is equipped for volumetric phallometry, that is, the apparatus measures penile blood volume change rather than penile circumference change. The volumetric method measures penile tumescence more accurately at low levels of response (Kuban, Barbaree, & Blanchard, 1999). A photograph and schematic drawing of the volumetric apparatus are given in Freund, Sedlacek, and Knob (1965). The major components include a glass cylinder that fits over the penis and an inflatable cuff that surrounds the base of the penis and isolates the air inside the cylinder from the outside atmosphere. A rubber tube attached to the cylinder leads to a pressure transducer, which converts air pressure changes into voltage output changes. Increases in penile volume compress the air inside the cylinder and thus produce an output signal from the transducer. The apparatus is calibrated so that known quantities of volume displacement in the cylinder—for example, 2 cc (cubic centimeters)—correspond to known changes in transducer voltage output. The apparatus is very sensitive and can reliably detect changes in penile blood volume much less than 1 cc. As measured by the Laboratory's equipment, full erection for the average patient corresponds to a blood volume increase of 20–25 cc. That is the blood volume increase for the part of the penis that projects into the glass cylinder, the only part that we can monitor.

The specific test used in this study has been described in substantial detail by Blanchard et al. (2001, 2007, 2009b). The

test stimuli were audiotaped narratives presented through headphones and accompanied by slides. There were seven categories of narratives, which described sexual interactions with prepubescent girls, pubescent girls, adult women, prepubescent boys, pubescent boys, and adult men, and also solitary, nonsexual activities (“neutral” stimuli). The accompanying slides showed nude models corresponding in age and sex to the topic of the narrative. Neutral narratives were accompanied by slides of landscapes. The test stimuli were presented as discrete trials, each 54 s in duration, with inter-trial intervals as long as necessary for penile blood volume to return to baseline. The full test consisted of four blocks of seven trials, with each block including one trial of each type in fixed pseudorandom order. All phallometric testing in this study was conducted by the same individual, a full-time staff member of the Laboratory. The time required to complete the test was usually about 1 h.

During the stimulus trials, penile blood volume change was sampled four times per second and recorded as a curve of blood volume change over time. For this study, the patient's response during a given trial was quantified as the greatest (positive or negative) change in blood volume from the moment of trial onset. These changes in blood volume were usually positive (i.e., increases) and were expressed in cc.² The data were further reduced to seven scores for each patient by averaging his four scores in each of the seven stimulus categories. These seven scores were taken as measures of the patient's erotic interest in adult women, pubescent girls, prepubescent girls, and so on.

Final Gating Criteria and Assignment to Groups

As stated earlier, the pool of potential subjects included 2,725 men. These were provisionally assigned to one of six groups according to their highest response on the phallometric test. Men who responded more to adult women than to any of the other six stimulus categories (including neutral stimuli) were classified as heterosexual teleiophiles; men who responded more to pubescent girls than to any of the other categories were classified as heterosexual hebephiles; men who responded more to prepubescent girls, as heterosexual pedophiles; men who responded more to prepubescent boys, as homosexual pedophiles; men who responded more to pubescent boys, as homosexual hebephiles; and men who responded more to adult

² We did not use the common, *z*-score conversion of the phallometric data for two reasons. The first was that *z*-scored data necessarily contain a large proportion of negative values (roughly about 50%). This would have forced us either to carry out some further transformation of the data to eliminate negative values, or else restrict ourselves to mathematical models that would work with negative values. The easiest course was simply not to use the *z*-score conversion. The second reason was that it seemed especially desirable, given all the other mathematics in this study, to express the results in an intuitive metric. It is more intuitive to relate penile erection to a change in physical volume (cubic centimeters) than to a difference in *z*-scores.

men than to any other category, as homosexual teleiophiles. There were 75 men who could not be classified according to their phallometric data, either because their highest penile response was to the neutral stimulus category (an outcome invariably associated with low responding) or because two different category scores were tied for first place. Their data were excluded from further analysis.

This left 2,650 subjects for further screening. In order to eliminate subjects whose phallometric data were relatively likely to be atypical for their group, we excluded men who completely denied any erotic interest in, or sexual experience with, persons resembling those to whom they responded most in the laboratory, and who furthermore had no known history of sexual offenses against such persons. Thus, we excluded a man from the study if his phallometric data put him in the homosexual teleiophile group, but his self-report indicated zero sexual interest in males over the age of 15, he had no known sexual offenses against males over the age of 15, and he reported no consenting sexual interactions, as an adult, with a male over the age of 17. We excluded a man from the study if his phallometric data put him in the homosexual hebephile group, but his self-report indicated zero sexual interest in males between the ages of 6 and 16, and he had no known sexual offenses against males between the ages of 6 and 16. Similarly, we excluded a man if his phallometric data put him in the homosexual pedophile group, but his self-report indicated zero sexual interest in males under the age of 15, and he had no known sexual offenses against males under the age of 15. The analogous exclusionary criteria were applied to the three heterosexual groups. This procedure identified 291 subjects whose phallometric group-assignment could not be supported by either their self-report or their known sexual history.

We conducted an ancillary data analysis to test our supposition that the phallometric data of the 291 excluded subjects would differ systematically from the phallometric data of the 2,359 remaining subjects. Since previous research has suggested that the reliability of phallometric tests is positively related to the amount that the subject responds to the stimuli (Lykins et al., 2010b), magnitude of response was an obvious choice of dependent variables. The amount of responding was quantified with a standard measure in the Kurt Freund Laboratory, the output index or OI (Freund, 1967). This is the average of the three greatest responses to any stimulus category except “neutral,” where penile response is expressed in cc of blood volume increase from the start of a trial.

Inspection of the data showed that the mean OI of the excluded subjects was roughly half that of the remaining subjects, 4.38 cc (SD = 6.01) vs. 8.10 cc (SD = 8.30). The reliability of this difference was examined in a 2×6 analysis of variance (ANOVA), in which the first factor was exclusion–nonexclusion status and the second factor was phallometric group-assignment. The results showed that the mean penile

response of the excluded subjects was significantly lower than that of the remaining subjects, $F(1, 2638) = 12.67$, $p = .0004$. There were also significant differences among phallometric groups, $F(5, 2638) = 3.88$, $p = .002$, but no interaction between phallometric group-assignment and exclusion–nonexclusion status, $F(5, 2638) < 1$. Thus, the results confirmed that men whose phallometric group-assignments could not be supported by either their self-reports or their known sexual histories did have phallometric results that were clearly atypical for their groups, at least regarding the one parameter we investigated.

The most common use of the OI in our laboratory is to identify patients whose penile blood volume changes during their phallometric testing stayed within the range typical of random blood volume fluctuations in nonaroused men. The phallometric test results of patients whose OI's are lower than 1.00 cc are routinely excluded from diagnostic consideration. We applied this criterion, in the present study, to the above-mentioned 2,359 remaining research candidates and excluded 81 more individuals on the basis of insufficient penile response. This left the 2,278 subjects used in the study: 1,066 heterosexual teleiophiles, 761 heterosexual hebephiles, 159 heterosexual pedophiles, 110 homosexual pedophiles, 86 homosexual hebephiles, and 96 homosexual teleiophiles.

Results

Figure 1 shows the phallometric response profiles of the six groups. By definition, the highest response of the heterosexual teleiophiles was to adult females, the highest response of the heterosexual hebephiles was to pubescent females, and so on. The ultimate aim of the following analysis can be understood as that of finding the best equation to predict each group's entire phallometric profile from its highest point alone.

The first step in data analysis was casting the summation and bipolar models in the form of competing equations, which are presented later. Each equation used the same input: the magnitude (in cc's) of the subject's penile response to his preferred stimulus category (i.e., the response that determined his group membership) and some measure of the “distance” from his preferred stimulus category to a given nonpreferred category. Each equation produced the same output: the predicted magnitude (again in cc's) of the subject's penile response to the given nonpreferred category.

This is more easily understood with a concrete example. Suppose that a subject's highest response was to pubescent girls (making him a heterosexual hebephile). The actual magnitude of this response was 12 cc of penile blood volume increase. The task of each equation would then be to predict the subject's response (in cc's) to adult women, prepubescent girls, prepubescent boys, pubescent boys, and adult men. Whichever equation made these predictions more accurately, more efficiently, or both would be considered the better model.

Table 1 Gender distances (G_i) from the preferred stimulus to the criterion stimuli for each group

Stimulus category	Group					
	Het teleios	Het hebes	Het pedos	Hom pedos	Hom hebes	Hom teleios
Adult women	0	0	0	1	1	1
Pubescent girls	0	0	0	1	1	1
Prepubescent girls	0	0	0	1	1	1
Prepubescent boys	1	1	1	0	0	0
Pubescent boys	1	1	1	0	0	0
Adult men	1	1	1	0	0	0

Note: We have used the term *criterion stimuli* rather than *nonpreferred stimuli* because we have included the preferred stimulus for each group and given its “distance” to itself

Het teleios heterosexual teleiophiles, *Het hebes* heterosexual hebephiles, *Het pedos* heterosexual pedophiles, *Hom pedos* homosexual pedophiles, *Hom hebes* homosexual hebephiles, *Hom teleios* homosexual teleiophiles

The concept of stimulus distance had to be operationalized differently for the two-dimensional summation model and the one-dimensional bipolar model. For the summation model, there was a gender distance and an age distance. If a given nonpreferred stimulus was the same gender as the preferred stimulus, then the gender distance was 0. If the nonpreferred stimulus was the opposite gender, then the distance was 1. Thus, for a heterosexual pedophile, the distance from the preferred stimulus (prepubescent girls) to adult women would be 0; the distance to prepubescent boys would be 1. Table 1 shows the gender distances, for each group, from their preferred stimulus to each nonpreferred stimulus.

Age distance in the summation model had three levels, which were quantified as 0, 1, and 2. If a given nonpreferred stimulus was in the same age-range as the preferred stimulus, then the age distance was 0. If the nonpreferred stimulus was one age-range away, the distance was 1; if the nonpreferred stimulus was two age-ranges away, the distance was 2. Thus, for a homosexual teleiophile, the age distance from the preferred stimulus (adult men) to adult women would be 0, the distance to pubescent girls would be 1, and the distance to prepubescent girls would be 2. Table 2 shows the age distances, for each group, from their preferred stimulus to each nonpreferred stimulus.

The bipolar model required only one distance measure, which will be referred to as morphological distance. Morphological distance was the number of steps between a subject's preferred stimulus category and any given nonpreferred category, on a hypothetical stimulus dimension ordered as follows: adult females, pubescent females, prepubescent females, prepubescent males, pubescent males, and adult males. (Only the relative positions of the stimulus categories matter; the dimension could just as easily be conceptualized as starting with adult males and ending with adult females.) Thus, for a heterosexual teleiophile, the morphological distance from the preferred stimulus (adult women) to pubescent girls would be 1, the distance to prepubescent girls would be 2, and the maximum distance—to adult men—would be 5. Table 3 shows the

morphological distances, for each group, from their preferred stimulus to each nonpreferred stimulus.

The tested equations were written as exponential equations with the design feature that the preferred stimulus—the stimulus category at zero distance from itself in either one- or two-dimensional space—would always “predict” its own value perfectly. The equation written to represent the summation model was

$$\hat{C}_i = P \times b_1^{G_i} \times b_2^{A_i},$$

where \hat{C}_i was the predicted magnitude (in cc's) of the subject's penile response to criterion stimulus i , P was the observed magnitude (in cc's) of the subject's response to his preferred stimulus (i.e., his highest response), G_i was the gender distance between the subject's preferred stimulus and criterion stimulus i (from Table 1), A_i was the age distance between the subject's preferred stimulus and criterion stimulus i (from Table 2), and b_1 and b_2 were parameters to be estimated from the data.³ The equation written to represent the bipolar model was

$$\hat{C}_i = P \times b^{M_i},$$

where M_i was the morphological distance between the subject's preferred stimulus and criterion stimulus i (from Table 3), and b was a parameter to be estimated from the data.⁴

In order to estimate the b parameters for these equations using all data from all subjects at once, we structured the data

³ The theoretical notion that differences in gender and in age are somehow additive is not readily apparent in the exponential equation because the equation does not include the arithmetic operation of addition. The concept of addition manifests when one solves the equation by taking the logarithm of both sides. That result (using a simplified notation for clarity) is: $\ln(C) = \ln(P) + (G \times \ln(b_1)) + (A \times \ln(b_2))$. Stated somewhat differently, the equation is additive at the logarithmic level.

⁴ Taking the logarithm of both sides of this equation yields the result $\ln(C) = \ln(P) + (M \times \ln(b))$.

Table 2 Age distances (A_i) from the preferred stimulus to the criterion stimuli for each group

Stimulus category	Group					
	Het teleios	Het hebes	Het pedos	Hom pedos	Hom hebes	Hom teleios
Adult women	0	1	2	2	1	0
Pubescent girls	1	0	1	1	0	1
Prepubescent girls	2	1	0	0	1	2
Prepubescent boys	2	1	0	0	1	2
Pubescent boys	1	0	1	1	0	1
Adult men	0	1	2	2	1	0

Note: We have used the term *criterion stimuli* rather than *nonpreferred stimuli* because we have included the preferred stimulus for each group and given its “distance” to itself

Het teleios heterosexual teleiophiles, *Het hebes* heterosexual hebephiles, *Het pedos* heterosexual pedophiles, *Hom pedos* homosexual pedophiles, *Hom hebes* homosexual hebephiles, *Hom teleios* homosexual teleiophiles

Table 3 Morphological distances (M_i) from the preferred stimulus to the criterion stimuli for each group

Stimulus category	Group					
	Het teleios	Het hebes	Het pedos	Hom pedos	Hom hebes	Hom teleios
Adult women	0	1	2	3	4	5
Pubescent girls	1	0	1	2	3	4
Prepubescent girls	2	1	0	1	2	3
Prepubescent boys	3	2	1	0	1	2
Pubescent boys	4	3	2	1	0	1
Adult men	5	4	3	2	1	0

Note: We have used the term *criterion stimuli* rather than *nonpreferred stimuli* because we have included the preferred stimulus for each group and given its “distance” to itself

Het teleios heterosexual teleiophiles, *Het hebes* heterosexual hebephiles, *Het pedos* heterosexual pedophiles, *Hom pedos* homosexual pedophiles, *Hom hebes* homosexual hebephiles, *Hom teleios* homosexual teleiophiles

file so that a “case” was defined by an observation rather than by a subject. There were six observations for each subject (his responses to adult women, pubescent girls, prepubescent girls, prepubescent boys, pubescent boys, and adult men); therefore the number of cases in the data file, 13,668, was six times greater than the number of subjects (2,278).

In addition to the observed response of one subject to one stimulus (the criterion stimulus, C), the record for each “case” included the same subject’s response to his preferred stimulus (P), the gender distance between the subject’s preferred stimulus and the criterion stimulus (G), the age distance between the subject’s preferred stimulus and the criterion stimulus (A), and the morphological distance between the subject’s preferred stimulus and the criterion stimulus (M). The criterion stimuli were obviously not statistically independent, since groups of six came from the same subject. The cases were treated as independent, however, in the analyses described next, because treating the data as a complex sample composed of clustered observations would have complicated the analyses greatly, and because the exact probability values automatically generated by these particular analyses were not really important.

The b parameters were estimated using nonlinear regression analysis (PASW—formerly SPSS—Version 17). For the summation model, the parameter estimate for gender distance was 0.291, 95% CI [0.282, 0.299], and the parameter estimate for age distance was 0.672, 95% CI [0.666, 0.679]. Thus, the regression equation corresponding to the summation model would be written

$$\hat{C}_i = P \times 0.291^{G_i} \times 0.672^{A_i}.$$

For the bipolar model, the parameter estimate for morphological distance was 0.633, 95% CI [0.628, 0.637]. Thus, the regression equation corresponding to the bipolar model would be written

$$\hat{C}_i = P \times 0.633^{M_i}.$$

Because these equations were so complicated to derive, it is important to stress how easy they are to use. Let us take, as an example, the raw data from one of the heterosexual pedophiles in the study. His greatest observed response, by definition, was to prepubescent girls; the actual magnitude of his response to this stimulus category was 12.62 cc. We will first calculate his predicted responses using the summation model and the dis-

tances presented in Tables 1 and 2. His predicted response to prepubescent girls was $12.62 \times 0.291^0 \times 0.672^0$, which equals $12.62 \times 1 \times 1$, or 12.62 cc. In other words, his predicted response to prepubescent girls was identical to his observed response to prepubescent girls, which is how the equation was designed to work. His predicted response to pubescent girls was $12.62 \times 0.291^0 \times 0.672^1$, which equals $12.62 \times 1 \times 0.672$, or 8.48 cc. His predicted response to adult men was $12.62 \times 0.291^1 \times 0.672^2$, which equals $12.62 \times 0.291 \times 0.452$, or 1.66 cc. His responses to adult women, prepubescent boys, and pubescent boys would be calculated in the same way.

The calculation of predicted responses using the bipolar model (in conjunction with the distances presented in Table 3) is even easier. Thus, the same subject's predicted response to adult men was 12.62×0.633^3 , which equals 12.62×0.253 , or 3.19 cc. This example also illustrates that the two models generate different predictions for nonpreferred stimuli. For this individual, the bipolar model predicted almost twice as much penile response to the stimulus category of adult men as did the summation model (3.19 vs. 1.66 cc).

Figure 2 shows the mean penile responses predicted by the summation model for all stimulus categories and for all groups. These data have been superimposed over the observed means already presented in Fig. 1. It is apparent from this figure that the predicted phallometric profiles for the six groups conform, at least roughly, to the observed phallometric profiles. A simple (i.e., zero-order) correlation coefficient was computed as a way of quantifying the level of agreement. This analysis used the previously described data file, which was structured so that a "case" was a set of observations rather than a subject. It would have been meaningless, however, to include the "cases" representing a subject's observed and predicted responses to his preferred stimulus category, because that pair of values was always identical. Since there was one such pair of values for each subject, their exclusion resulted in a sample of $13,668 - 2,278 = 11,390$ cases. The correlation between the observed and predicted penile responses to the nonpreferred stimulus categories was $r = .755$.

Figure 3 shows the mean penile responses predicted by the bipolar model for all stimulus categories and for all groups. As in Fig. 2, the data have been superimposed over the observed means presented in Fig. 1. Visual inspection suggested that the agreement between the predicted and observed phallometric profiles was, considering all groups, at least as good as that obtained using the summation model. That suggestion was supported by the correlation analysis; the correlation between the observed and predicted responses to the nonpreferred stimulus categories was $r = .778$.

The next phase of data analysis focused on more formal comparisons of goodness of fit. In this phase, we sought to determine which model made smaller errors in predicting a

subject's responses to his nonpreferred stimulus categories from his response to his preferred stimulus category. To this end, we devised a special measure of prediction error, which was computed separately for the summation and bipolar models. The calculation of this measure is described below.

First, we computed the difference between the subject's observed response to each of his nonpreferred stimulus categories and the response predicted by either the summation or the bipolar equation presented above. These quantities were, of course, the same as the unstandardized residuals from the nonlinear regression analyses. These residuals were computed only for the subject's five nonpreferred stimulus categories, because the residual for the preferred stimulus category was always equal to zero. Second, we took the absolute value of each residual. Third, we computed the average of the five absolute values for each subject. We called this quantity the *profile discrepancy index*. Smaller values on the profile discrepancy index meant a better fit between the subject's observed phallometric profile and the profile predicted by an equation.

For this part of the analysis, we returned to a standard file structure, that is, one "case" or record for each subject. Thus, in what follows, sample sizes refer to numbers of subjects, not to numbers of observations. The sample is no longer complex, and the p values, which are now important for the interpretation of the results, can be taken at face value.

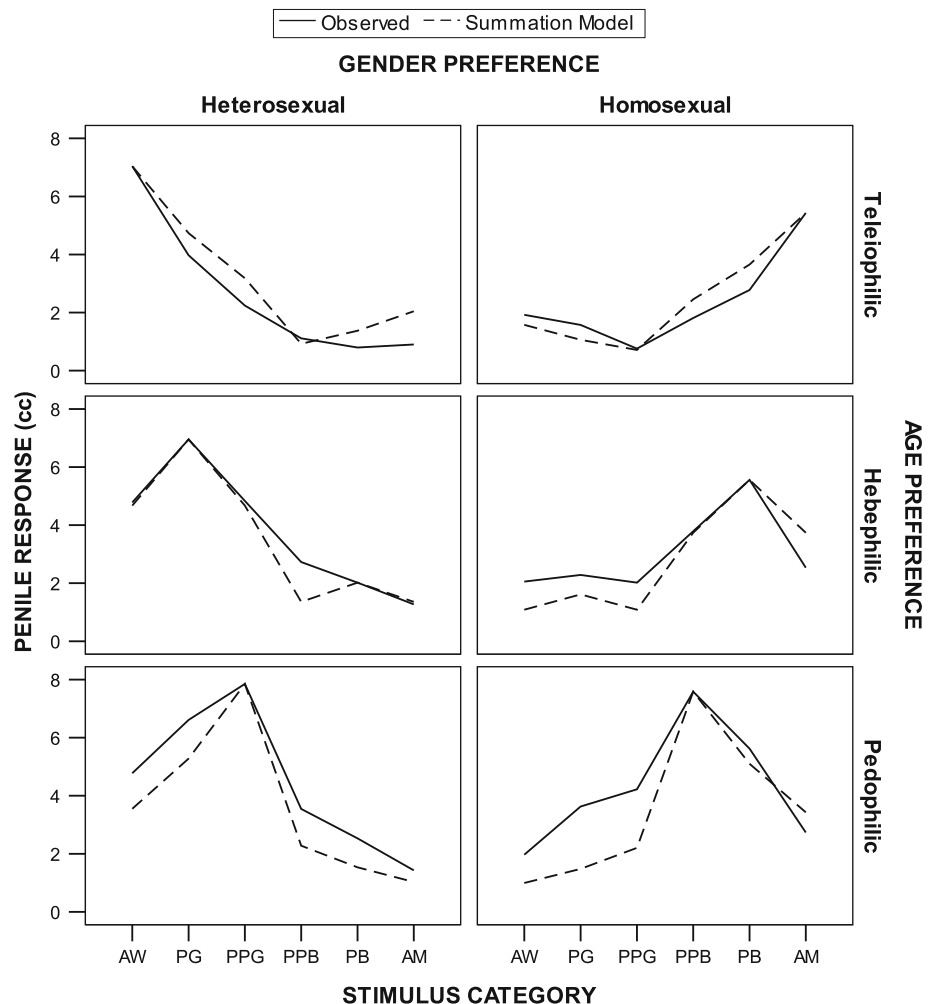
We used a mixed-design ANOVA to compare the mean profile discrepancy index for the summation and bipolar models. The between-subjects variable was group assignment. Since there were six groups, this variable had six levels. The within-subjects variable was mathematical model (summation or bipolar), which had two levels. There was a significant main effect for group, $F(5, 2272) = 3.19$, $p = .007$, and a significant group \times model interaction, $F(5, 2272) = 11.43$, $p \ll 10^{-6}$. However, the most important effect for our purposes—the main effect for mathematical model—did not reach statistical significance, $F(1, 2272) = 2.92$. In other words, we did not find a difference, within the sample as a whole, between the profile discrepancy index associated with the summation model and the profile discrepancy index associated with the bipolar model.

In the comparison of non-nested mathematical models, a draw is usually considered a win for the model with fewer parameters, which in this case was the bipolar model (1 parameter). We decided, however, to try not to choose between models on purely technical grounds. The strategy we pursued instead was to upgrade both models (by adding one additional parameter to each), and to investigate whether a more clear-cut difference emerged when we compared the upgraded models.

The selection of additional parameters was fairly obvious in both cases. For the summation model, we simply added a parameter for the interaction of stimulus-gender and stimulus-age. Thus,

Fig. 2 Observed phallometric profiles and profiles predicted by the summation model.

Abbreviations for stimulus categories: *AW* adult women, *PG* pubescent girls, *PPG* prepubescent girls, *PPB* prepubescent boys, *PB* pubescent boys, *AM* adult men



$$\hat{C}_i = P \times b_1^{G_i} \times b_2^{A_i} \times b_3^{(G_i \times A_i)}.$$

The selection of the additional parameter for the bipolar model was prompted by both a priori and empirical considerations. In the arrays of morphological distances presented in Table 3, gender is completely invisible. The distance between pubescent girls and prepubescent girls (one unit) is the same as the distance between prepubescent girls and prepubescent boys. Similarly, the distance between prepubescent girls and prepubescent boys is the same as the distance between prepubescent boys and pubescent boys. One might readily question whether the distance between prepubescent girls and prepubescent boys should be somewhat larger than the simple unit distance employed elsewhere. In other words, crossing the gender line might count a little extra.

That this potential problem is an actual problem is illustrated by the data shown in Fig. 3. The original equation written to represent the bipolar model implies that a heterosexual pedophile's response to prepubescent boys should be equal to his response to pubescent girls, because they are equidistant from his preferred category of prepubescent girls. Similarly, a homosexual pedophile's response to prepubescent girls should

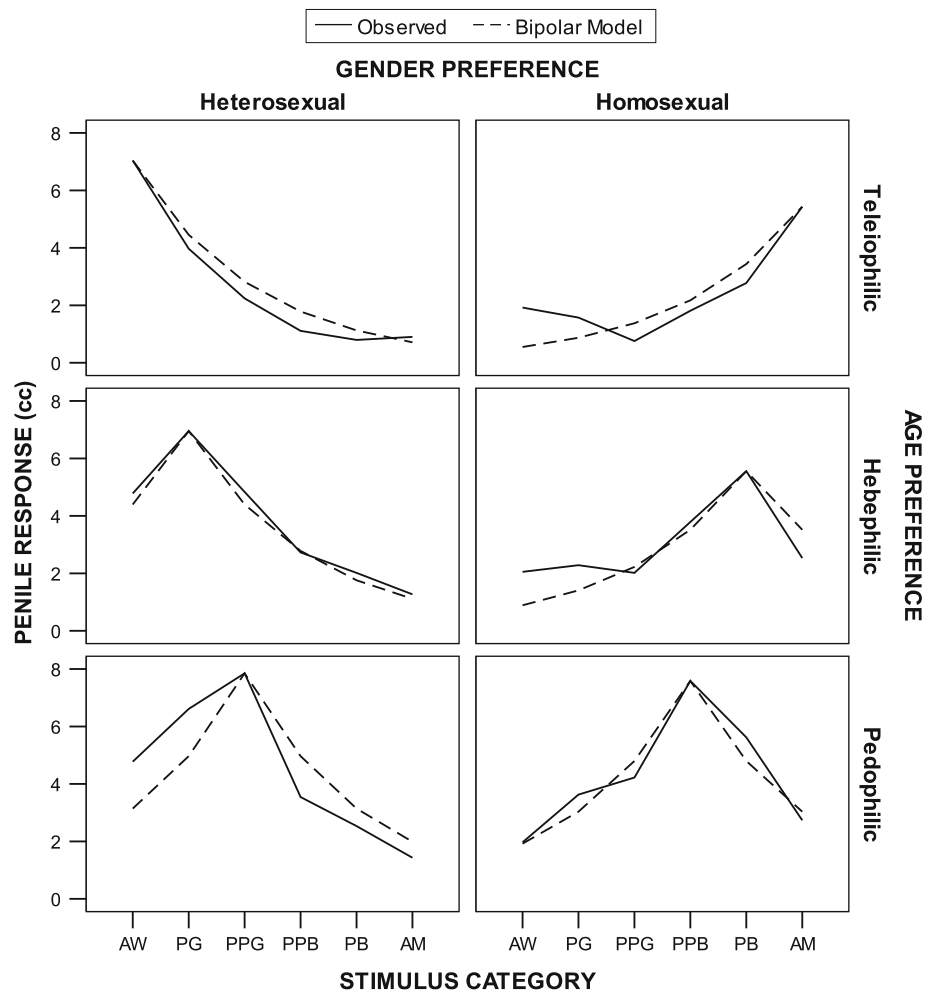
be equal to his response to pubescent boys. The bottom panels of Fig. 3 indicate that neither is the case. The heterosexual pedophiles responded more to pubescent girls than to prepubescent boys, and the homosexual pedophiles responded more to pubescent boys than to prepubescent girls. *T*-tests for pairs showed that these differences were significant, both for the heterosexual pedophiles, $t(158) = 7.89, p \ll 10^{-6}$, and for the homosexual pedophiles, $t(109) = 3.19, p = .002$. It is therefore clear that the morphological distance between prepubescent girls and prepubescent boys should be represented, in our study, by a number larger than 1. But how much larger?

Fortunately, it was possible to estimate the optimal increment to the prepubescent girl–prepubescent boy distance by adding a single parameter to the original equation for the bipolar model. This parameter therefore became the single improvement allocated for that equation. The new equation was written as follows:

$$\hat{C}_i = P \times b_1^{(M_i + (b_2 \times G_i))}.$$

Since G_i (Table 1) is either 0 or 1, the effect of this parameter would be to add some constant, b_2 , to every morphological

Fig. 3 Observed phallometric profiles and profiles predicted by the bipolar model. Abbreviations for stimulus categories: *AW* adult women, *PG* pubescent girls, *PPG* prepubescent girls, *PPB* prepubescent boys, *PB* pubescent boys, *AM* adult men



distance M_i (Table 3) that crosses the gender line. Thus, if b_2 were determined to be 0.7, for example, that would mean that the distance between prepubescent girls and prepubescent boys should be 1.7 rather than 1, and the column of distances in Table 3 for the heterosexual telephiles would read: 0, 1, 2, 3.7, 4.7, 5.7. The column for heterosexual pedophiles, to give a second example, would read 2, 1, 0, 1.7, 2.7, 3.7, and the stimulus distances from prepubescent girls to pubescent girls and to prepubescent boys would no longer be equal.

The parameters for the two revised equations were estimated using nonlinear regression analysis, as described before. For the summation model, the parameter estimate for gender distance was 0.248, 95% CI [0.238, 0.258], the parameter estimate for age distance was 0.661, 95% CI [0.654, 0.668], and the parameter estimate for the interaction was 1.352, 95% CI [1.296, 1.407]. Thus, the revised regression equation corresponding to the summation model would be written

$$\hat{C}_i = P \times 0.248^{G_i} \times 0.661^{A_i} \times 1.352^{(G_i \times A_i)}.$$

For the bipolar model, the parameter estimate for morphological distance was 0.661, 95% CI [0.655, 0.667], and

the gender-crossing correction factor was 0.502, 95% CI [0.413, 0.591]. Thus, the revised regression equation corresponding to the bipolar model would be written

$$\hat{C}_i = P \times 0.661^{(M_i + (0.502 \times G_i))}.$$

Note that the b_2 parameter estimate implies that the perceived morphological distance between prepubescent girls and prepubescent boys is close to 1.5 times as large as the distances between the other stimulus categories.

For both revised models, the correlation between the observed and predicted penile responses to the nonpreferred stimulus categories was very close to the value obtained with the original model. These correlations increased minutely to .760 and .780 for the summation and bipolar models, respectively.

To see whether the added parameters would actually lower the profile discrepancy index, we compared the original and expanded summation equations on that variable, and we compared the original and expanded bipolar equations on that variable. The comparisons were carried out on the full sample, ignoring group, and they used the t -test for pairs. The mean profile discrepancy index for the summation model

actually went up, 1.818 cc (SD = 2.03) vs. 1.824 cc (SD = 2.05), $t(2277) = -2.71$, $p = .007$; in other words, the fit—at least, as measured by the profile discrepancy index—actually got worse with the interaction term added to the summation model. In contrast, the mean profile discrepancy index for the bipolar model went down, 1.775 cc (SD = 1.98) vs. 1.760 cc (SD = 1.96), $t(2277) = 4.23$, $p < .0001$, indicating that increasing the relative stimulus distance between prepubescent girls and prepubescent boys improved that model. Figure 4 shows the mean penile responses predicted by the revised bipolar model for all stimulus categories and for all groups, superimposed over the observed means. Because the revised summation model was not considered further in this study, for reasons explained below, we have not presented graphic data for it.

The results of the foregoing analyses indicated that there was little point in proceeding with the plan to re-run the ANOVA comparing the mean profile discrepancy indices for the summation and bipolar models using indices based on the revised equation for each model. It would have been difficult to tell whether a different outcome was caused by the improvement in the bipolar model, the worsening of the summation model, or both. We therefore decided to re-run this comparison using the better version of each model: the original version of the summation model and the revised version of the bipolar model. As before, we used a mixed-design ANOVA to compare the mean profile discrepancy index for the two models. The between-subjects variable was group assignment, and the within-subjects variable was mathematical model (original summation or revised bipolar). This time there was a small but statistically significant main effect for mathematical model, $F(1, 2272) = 11.87$, $p = .001$, partial $\eta^2 = .005$. Overall, the profile discrepancy index associated with the bipolar model was lower than the profile discrepancy index associated with the summation model; the grand means were 1.76 cc (SD = 1.96) and 1.82 cc (SD = 2.03), respectively. This indicates that the bipolar model provided a better fit to the observed data. There was also a significant main effect for group, $F(5, 2272) = 2.93$, $p = .01$, and a significant group \times mathematical model interaction, $F(5, 2272) = 9.15$, $p < 10^{-6}$.

Because of the significant interaction effect, it was necessary to compare the profile discrepancy indices for the summation and bipolar models for each group separately. The results are shown in Table 4. The profile discrepancy index produced by the bipolar model was significantly lower for the heterosexual teleiophiles, the heterosexual hebephiles, and the homosexual pedophiles, indicating that the bipolar model provided a better fit for those groups. For the remaining three groups, it was not possible to demonstrate any superiority of one model over the other.

The final analysis took an ancillary, indirect approach to assessing the predictive accuracy of the two models, an approach that was not based on comparing the residuals generated by the summation and bipolar models. This approach was

based on a derived variable suggested by Fig. 1. Visual inspection of Fig. 1 suggested that the six groups might differ in regard to their mean response to the five nonpreferred stimulus categories. This was tested empirically. We calculated, for each subject, a single score equal to the average of his observed responses to his five nonpreferred stimulus categories.⁵ The unit of measurement for this variable was cc's of penile blood volume change. The group means for this variable are presented in Fig. 5. The group means were analyzed in a one-way ANOVA, which showed that they were significantly different, $F(5, 2272) = 18.44$, $p < 10^{-6}$. A Scheffé multiple-range test at the $p < .01$ level showed that the two pedophilic groups differed significantly from the two teleiophilic groups. The two hebephilic groups, whose means fell between those of the pedophiles and the teleiophiles, did not differ significantly from either.

The foregoing analysis laid the foundation for our second test of the relative predictive accuracy of the two models. In addition to the average of the observed responses to the five nonpreferred stimulus categories, we calculated, for each subject, the average of the five responses predicted by the (original) summation model, and the average of the five responses predicted by the revised bipolar model. Figure 6 shows the group means for all three of these quantities.

Figure 6 strongly suggests that the mean observed response to the nonpreferred stimulus categories was better predicted by the bipolar model than by the summation model. This is particularly true for the pedophilic groups, whose response to their nonpreferred stimulus categories was notably underpredicted by the summation model.

The data shown in Fig. 6 were formally analyzed in a mixed-design ANOVA. The dependent variable was the subject's averaged penile response to his five nonpreferred stimulus categories. The between-subjects variable was group assignment. Since there were six groups, this variable had six levels. The within-subjects variable was the type of data. This variable had three levels: observed response, response predicted by the summation model, and response predicted by the bipolar model.

The main effect for group showed, as expected, that differences among the six groups were statistically significant, $F(5, 2272) = 10.94$, $p < 10^{-6}$. The main effect for type of data showed that there were differences among the observed and predicted responses, $F(2, 4544) = 13.02$, $p < .0001$. There was also a significant interaction between group and type of data, $F(10, 4544) = 24.48$, $p < 10^{-6}$.

The key results from this ANOVA concerned specific contrasts. Tests of within-subjects contrasts showed that the observed

⁵ This score is roughly proportional to the “area under the curve” of the subject's phallometric profile.

Fig. 4 Observed phallometric profiles and profiles predicted by the revised bipolar model. Abbreviations for stimulus categories: *AW* adult women, *PG* pubescent girls, *PPG* prepubescent girls, *PPB* prepubescent boys, *PB* pubescent boys, *AM* adult men

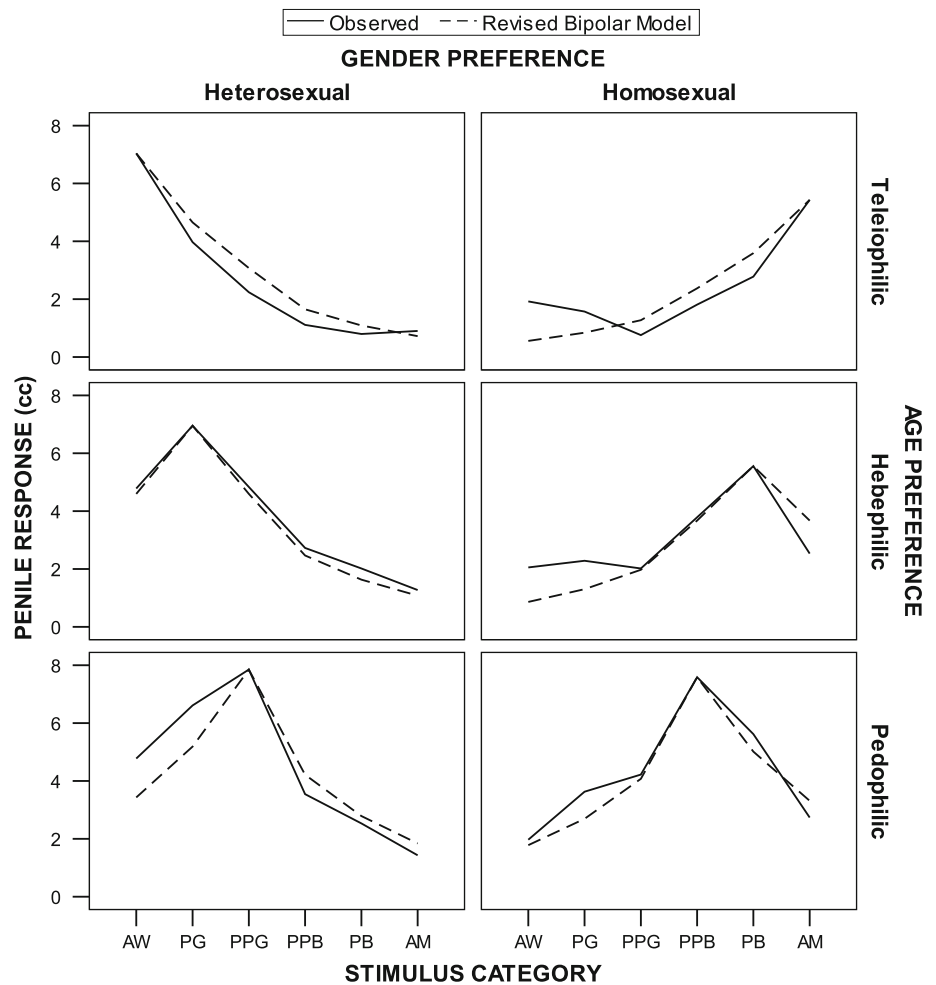


Table 4 Pairwise comparisons of the profile discrepancy index for the summation and bipolar models

Summation model		Bipolar model		<i>F</i>	df	<i>p</i>
<i>M</i>	SD	<i>M</i>	SD			
Heterosexual teleiophiles						
1.86	1.95	1.79	1.88	33.08	1, 2272	≪.0001
Heterosexual hebephiles						
1.68	1.97	1.63	1.94	10.15	1, 2272	.001
Heterosexual pedophiles						
2.16	2.73	2.20	2.58	1.82	1, 2272	n.s.
Homosexual pedophiles						
2.24	2.69	1.98	2.36	42.78	1, 2272	≪.0001
Homosexual hebephiles						
1.52	1.30	1.52	1.29	0.01	1, 2272	n.s.
Homosexual teleiophiles						
1.70	1.65	1.76	1.65	1.62	1, 2272	n.s.

Note: The unit of measure for the means (*M*) and standard deviations (SD) of the profile discrepancy index is cubic centimeters (cc) of penile blood volume

Fig. 5 Average of observed responses to the five nonpreferred stimulus categories. Abbreviations for groups: *Het* heterosexual, *Hom* homosexual, *Teleios* teleiophiles, *Hebes* hebephiles, *Pedos* pedophiles

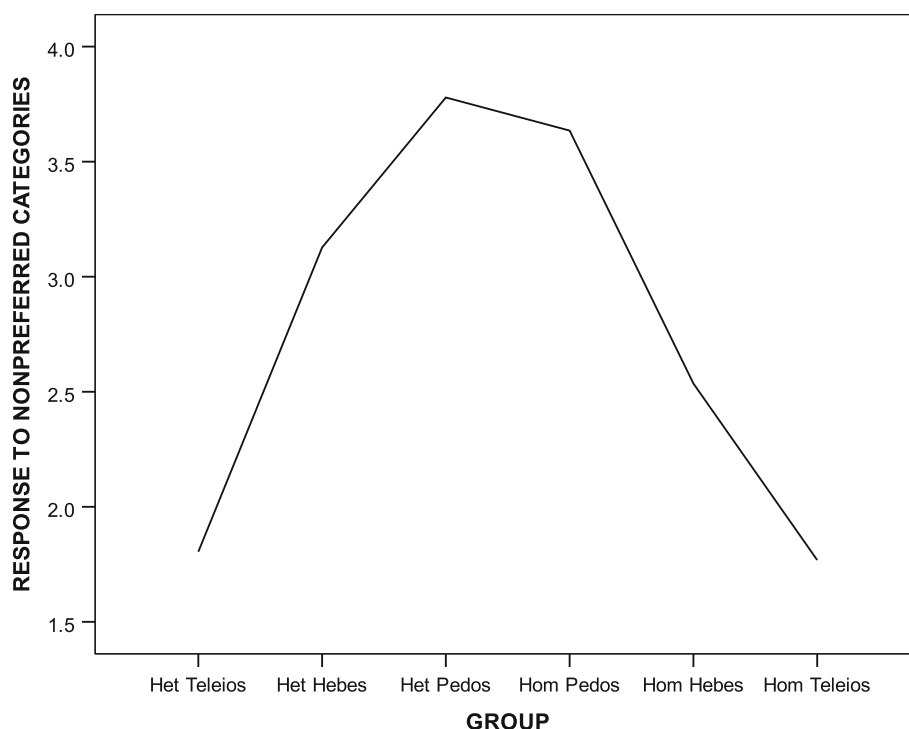
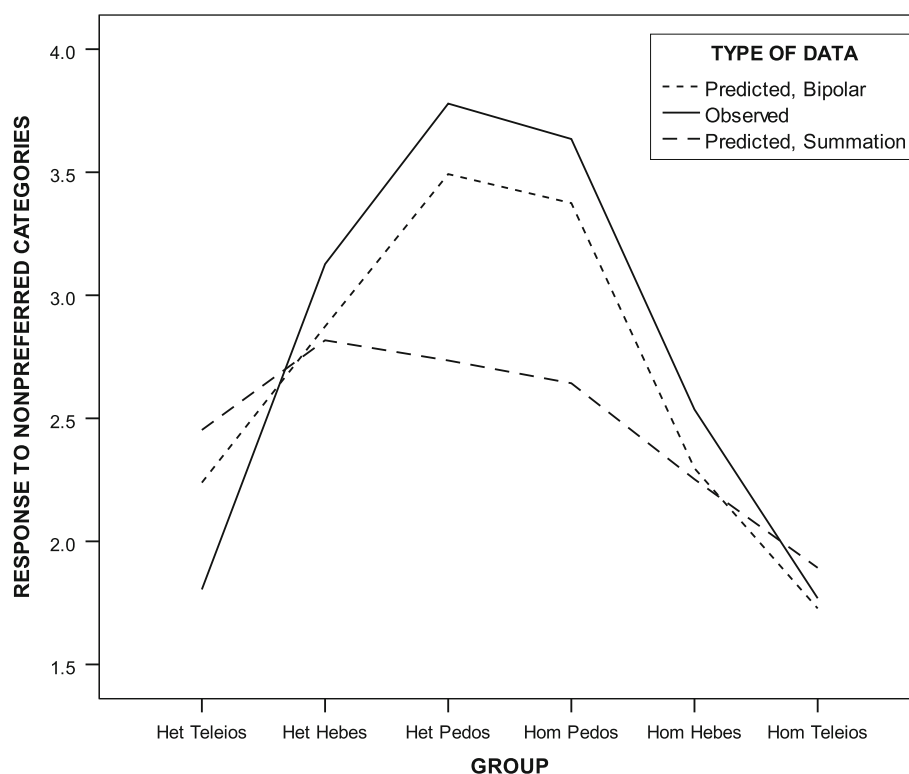


Fig. 6 Average of the observed responses to the five nonpreferred stimulus categories, average of the five responses predicted by the summation model, and the average of the five responses predicted by the revised bipolar model. Abbreviations for groups: *Het* heterosexual, *Hom* homosexual, *Teleios* teleiophiles, *Hebes* hebephiles, *Pedos* pedophiles



responses did not differ from those predicted by the bipolar model, $F(1, 2272) = 2.11$. The observed responses did, however, differ significantly from those predicted by the summation model, $F(1, 2272) = 16.61$, $p < .0001$. Thus, the bipolar model again appeared superior to the summation model.

Discussion

This study compared two psychophysiological models of alloerotic responding (sexual responding to other people) in men. The first model was based on the notion that men respond to a potential sexual object as a compound stimulus made up of an

age component and a gender component. The second model was based on the notion that men respond to a potential sexual object as a gestalt, which they evaluate in terms of global similarity to other potential sexual objects. The analytic strategy was to compare the accuracy of these models in predicting a man's penile response to each of his less arousing (nonpreferred) stimulus categories from his response to his most arousing (preferred) stimulus category. Both models based their predictions on the degree of dissimilarity between the preferred stimulus category and a given nonpreferred stimulus category, but each model used its own measure of dissimilarity. According to the first model, penile response should vary inversely as the sum of stimulus differences on separate dimensions of age and gender. It was therefore called the summation model. According to the second model, penile response should vary inversely as the distance between stimulus categories on a single, bipolar dimension of morphological similarity—a dimension on which children are located near the middle, and adult men and women are located at opposite ends. The second model was accordingly called the bipolar model. The results, subject to the qualifications discussed later, favored the bipolar model. This implies that men respond sexually as if they perceive other humans as points along a single, bipolar dimension of morphological similarity—a stimulus dimension in which adult men are located at one end, prepubescent children are located near the midpoint, and adult women are located at the opposite end.

Comparing the bipolar and summation models required that the stimulus categories be ordered exactly as we have them on the *X*-axes of Figs. 1, 2, 3, and 4. It is, in fact, difficult to imagine how the bipolar equation, especially in its original, one-parameter form, could possibly have worked if these categories were arranged in a different order. This definition of the stimulus continuum also makes it easy to visualize and conceptualize sexual orientations as a series of overlapping stimulus generalization gradients. This is exemplified in Fig. 7, which shows the phallometric profiles predicted by the revised bipolar equation for the six types of men examined in the present research.

It is instructive to compare the concept of sexual orientations as a series of overlapping generalization gradients with the best known previous visualizable concept of sexual orientations, the so-called “Kinsey Scale” (Kinsey, Pomeroy, & Martin, 1948). The Kinsey Scale is, of course, quite different from the generalization gradient model in that it is concerned solely with erotic gender-preference, not with erotic age-preference. Furthermore, it is implicit in most applications or discussion of the Kinsey Scale that it is intended for teleiophiles; in other words, it is concerned with relative attraction to male versus female adults. That is not the point of difference we wish to discuss here, however.

The Kinsey continuum is essentially an ordering of *responses*. A person rated as a “3” (the mid-point on the Kinsey Scale) is someone who is equally attracted to physically normal men and women, not someone who is preferentially attracted to persons with ambiguous genitalia (intersexes) or with some other sexually intermediate body type. In contrast, our model's continuum is essentially an ordering of *stimuli*. The subject's responses are located on a different axis. Thus, the two models are fundamentally different in conceptualization, not only in scope.

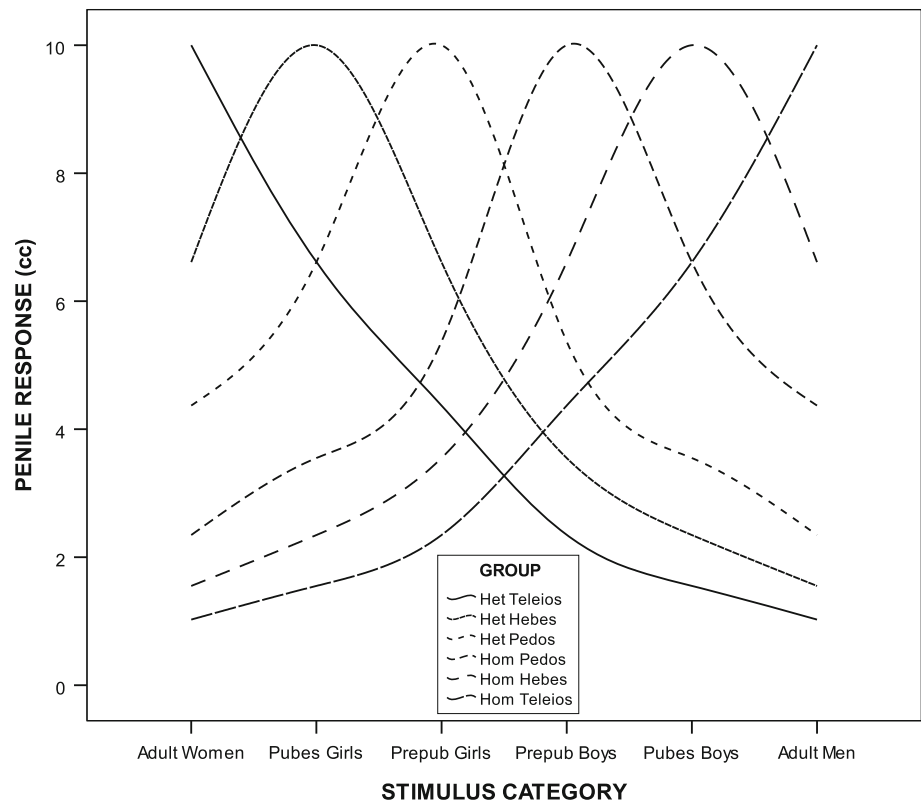
The bipolar model implies that bisexuality should be relatively common in pedophiles. The available data do, in fact, indicate that bisexuality is more common in pedophiles (e.g., Blanchard et al., 1999; Carlstedt et al., 2009) than in teleiophiles (e.g., Laumann, Gagnon, Michael, & Michaels, 1994, p. 311, Table 8.3A). Thus, data based on sexual partners and victims are compatible with at least one conclusion implied by our psychophysiological findings.

Bisexuality in male teleiophiles raises a special problem for the bipolar model. The phallometric responses of bisexual teleiophiles could be plotted against an *X*-axis like that used in Figs. 1, 2, 3, 4, and 7. The phallometric profile should be V-shaped or U-shaped. Such a profile could not, however, be described by the bipolar model equation; it would require the summation model equation. This would be a limitation for the bipolar model; how serious a limitation depends on the proportion of self-reported bisexual teleiophiles who actually produce V-shaped or U-shaped phallometric profiles.

There is at least one other class of men whose existence is unquestioned, and whose behavior cannot be explained by the bipolar model. Those are the *gynandromorphophiles*, men who are sexually attracted to those individuals colloquially known as “she-males.” She-males are biological males who have partially feminized their bodies with estrogenic hormones or breast implants but have not undergone surgical modification of the genitals, thus creating the appearance of a woman with a penis (Blanchard & Collins, 1993). She-males are commonly employed in sex work or the adult entertainment industry. Although persons with disorders of sex development (i.e., intersexes) occur in nature, voluptuous women with large penises do not. Thus, the stimulus category of greatest, or at least substantial, interest to gynandromorphophiles does not occur on the continuum presupposed by the bipolar model.

Neither the bipolar model nor the (provisionally rejected) summation model says anything about etiology. These models are not concerned with the routes by which men come to be pedophiles, hebephiles, or teleiophiles; they are concerned with men's behavior once they get there. Figure 1, even without any statistical analysis, shows that all the groups in the study demonstrated stimulus generalization in the laboratory, and thus behaved similarly in this particular regard. However,

Fig. 7 Phallometric profiles predicted by the revised bipolar model for heterosexual and homosexual pedophiles, hebephiles, and teleiophiles. All profiles were calculated assuming a penile response of 10 cc to the preferred stimulus category. Abbreviations for groups: *Het* heterosexual, *Hom* homosexual, *Teleios* teleiophiles, *Hebes* hebephiles, *Pedos* pedophiles. Abbreviations for stimulus categories: *Pubes* pubescent, *Prepub* prepubescent



one might ask—in the spirit of pursuing between-groups differences that might lead to etiological hypotheses—did all the groups generalize to the same extent?

This question is not as simple to answer as it might seem. When we quantified stimulus generalization as the average of the subject's observed responses to his five nonpreferred stimulus categories, we found that the two pedophilic groups responded significantly more than the two teleiophilic groups. This might be interpreted to mean that the pedophiles generalized more than the teleiophiles (or, stated differently, that pedophiles were less discriminating in their sexual responses). There are two ways that this result could come about, however: (1) There is something different about pedophiles' sexual reactions, and (2) there is something different about pedophiles' preferred stimulus categories. The latter possibility arises because pedophiles' preferred stimulus categories are in the center of the stimulus continuum; pedophiles can generalize to the right and to the left, so to speak. Teleiophiles' preferred stimulus categories are at the ends of the continuum; thus they can generalize only to one side.

It turned out that the bipolar equation (in its original or revised form) predicts or even *requires* that the pedophilic group have the highest response to their nonpreferred stimulus categories and that the teleiophilic group have the lowest. Thus, a mathematical model that describes pedophiles and teleiophiles as behaving in exactly the same way also predicts that pedophiles will seem to show more stimulus generalization (or less

discrimination), if that is quantified as their averaged response to all of their nonpreferred stimulus categories.

In summary, then, we found no evidence that pedophiles respond either more or less discriminately than hebephiles or teleiophiles. Pedophiles respond more throughout the test, not because they discriminate less between their preferred stimulus category and its closely neighboring categories, but because their preferred stimulus category has more closely neighboring categories.

Another goal of the study, as stated in the Introduction, was to investigate whether the same model worked best for each of our six subgroups of men. We were able to show that the bipolar model provided a better fit for the heterosexual teleiophiles, the heterosexual hebephiles, and the homosexual pedophiles. For the remaining three groups, it was not possible to demonstrate any superiority of one model over the other. Since there was no particular pattern to the occurrence of significant and nonsignificant results, a conservative interpretation is that the inconsistency at the level of separate groups was caused by one or more weaknesses of the study.

There were two kinds of potential weaknesses in the study: those specific to our particular subjects and stimulus materials, and those inherent in our basic approach. The specific weaknesses will be discussed first. Most of the subjects in our study were sex offenders, and most of those were probably motivated to produce the most "normal-looking" phallometric profiles they could. We made an (apparently successful)

attempt to eliminate subjects whose phallometric data were relatively likely to be atypical for their group. We did not, however, limit the research to men who appeared fully candid about their erotic preferences, because that would have reduced the sample more than desirable for this study (see Blanchard et al., 2009b). It is therefore likely that the present data retained some social desirability response bias, but it would have been difficult to determine how much and in which groups.

The second specific weakness concerns the phallometric stimuli. Our theoretical models were cast solely in terms of morphological differences between males and females, adults and children. Our multimedia phallometric stimuli, however, included audiotaped narratives that described sexual interactions with prepubescent girls, pubescent girls, adult women, prepubescent boys, pubescent boys, or adult men. These narratives were presented simultaneously with slides that showed nude models corresponding in age and gender to the topic of the narrative. The narratives are important for clinical diagnosis because they increase the magnitude of response (Lykins et al., 2010b). In the context of the present research, however, they represent a potential source of stimulus contamination. It is, more to the point, impossible to know whether they influenced the responding of some groups more than others.

The potential inherent weakness concerns the symmetry of the distance arrays, particularly the morphological distances shown in Table 3. The problem is most easily illustrated by example. Are prepubescent girls just as different from adult men as prepubescent boys are from adult women? The difference in height between prepubescent girls and adult men is greater than the difference in height between prepubescent boys and adult women, and the difference in hirsuteness is also greater. These are just two examples; others could be advanced. The reliance on symmetrical distances is another potential source of error in our mathematical models, and one that could have differential effects on the various groups.

The third goal of the research was to determine whether it would be possible to express the summation and bipolar dimension models in the form of competing equations that could predict all the points on a man's phallometric response profile solely from the magnitude of his highest response—regardless of whether his highest response is to males or females, to children, pubescents, or adults. Clearly this was possible, at least for aggregate data. What is remarkable is that it was possible to do a credible job of approximating the observed data using an equation (the original bipolar model) that included only a single estimated parameter. The addition of a second parameter to the bipolar equation did produce some statistical improvement, although the difference was not visually striking (cf. Figs. 3, 4). Because the second parameter has potential significance for our basic theoretical question, it is worthwhile devoting some attention to it.

The addition of the second estimated parameter to the bipolar model had the same effect as adding 0.5 units, in Table 3, to the distance between a subject's preferred stimulus category and any nonpreferred stimulus category that differed in gender from the preferred category. Thus, for example, the distances, for heterosexual teleiophiles, from adult women to pubescent girls and prepubescent girls remained the same, at 1 and 2 units, respectively; but the distances to prepubescent boys, pubescent boys, and adult men increased from 3, 4, and 5 units to 3.5, 4.5, and 5.5 units.

There are two ways of looking at this. The first is that the revision fully preserved the concept of a single continuum of morphological similarity and merely adjusted the relative distances of stimulus categories along that continuum. There was, after all, no empirical basis for the original assignment of equal, integer values to the distances between all neighboring categories. It stands to reason that at least one of these assigned distances would differ notably from its ideal, platonic counterpart, and a likely candidate for requiring correction was the distance between prepubescent girls and prepubescent boys.

The second way of looking at the revised bipolar model is that the addition of the second parameter is a mathematical acknowledgment that men do indeed perceive gender as a separate stimulus feature in potential erotic objects. On this view, the revised equation can no longer be claimed to represent a pure unidimensional model. It represents a kind of hybrid, which incorporates an important element of the summation hypothesis. There is, at present, no way to decide between these views, and the "meaning" of the second parameter must remain ambiguous.

There is one final point to be made regarding the bipolar model. Both the original and revised versions of the bipolar equation predict that heterosexual teleiophiles will respond more to prepubescent boys than to pubescent boys, and more to pubescent boys than to adult men (see Fig. 7). They make the analogous prediction for homosexual teleiophiles. Previous research has found, however, that heterosexual teleiophiles respond equally and minimally (or undetectably) to males of all ages, and that homosexual teleiophiles respond equally and minimally to females of all ages (e.g., Freund et al., 1973).

Does this mean that the bipolar equation fails to hold for stimulus categories beyond a certain distance from the preferred category—that beyond a certain distance, all stimulus categories produce the same minimal degree of penile response (or no penile response)? Not necessarily. It is possible that the "flat" segments of the phallometric profiles of heterosexual and homosexual teleiophiles reflect a floor effect that equalizes observed responses to nonpreferred stimulus categories among low and moderately reactive subjects. This possibility is illustrated by data from Blanchard et al. (2009a). That study included a group of heterosexual teleiophiles with very high responses. Their responses to prepubescent boys,

pubescent boys, and adult men showed a downward trend that resembles the predicted relations shown in Fig. 7.

The notion that latent differences in response to nonpreferred stimulus categories may become detectable when subjects' reactivity is increased is bolstered by results obtained by Chivers, Seto, and Blanchard (2007). Their subjects included heterosexual and homosexual male volunteers between the ages of 18 and 40 years. The phallometric stimuli included videotapes of men and women engaging in same-sex intercourse, engaging in solitary masturbation, or performing nude exercise (no sexual activity). These videotapes were much stronger stimuli than the slides plus audiotapes used in the present study. Chivers et al. (2007) found that heterosexual men had significant increases in penile response to videotapes of two men having sexual intercourse, but not to solitary men masturbating themselves or to men exercising in the nude. Similarly, homosexual men showed an increase in penile response to female–female intercourse that approached statistical significance, but not to female masturbation or nude exercise. The studies by Blanchard et al. (2009a) and Chivers et al. (2007) suggest that the bipolar model might predict adequately throughout the full range of nonpreferred stimulus categories when laboratory conditions are optimal.

Summary and Conclusions

Sexual orientations are reflected in phallometric data as a series of overlapping stimulus generalization gradients. The shapes of these gradients can be described, for a very wide variety of men, with the same, single-parameter, exponential equation. This equation can be improved by the addition of a second parameter, although the difference is not great. The equation is compatible with the hypothesis that men respond to a potential sexual object as a gestalt, not as a compound stimulus made up of an age component and a separate gender component. There may be certain classes of men, including bisexual teleiophiles and gynandromorphophiles, for whom the same basic equation does not hold, and whose behavior would be better described by a different model.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Blanchard, R., & Collins, P. I. (1993). Men with sexual interest in transvestites, transsexuals, and she-males. *Journal of Nervous and Mental Disease*, 181, 570–575.
- Blanchard, R., Klassen, P., Dickey, R., Kuban, M. E., & Blak, T. (2001). Sensitivity and specificity of the phallometric test for pedophilia in nonadmitting sex offenders. *Psychological Assessment*, 13, 118–126.
- Blanchard, R., Kolla, N. J., Cantor, J. M., Klassen, P. E., Dickey, R., Kuban, M. E., et al. (2007). IQ, handedness, and pedophilia in adult male patients stratified by referral source. *Sexual Abuse: A Journal of Research and Treatment*, 19, 285–309.
- Blanchard, R., Kuban, M. E., Blak, T., Cantor, J. M., Klassen, P. E., & Dickey, R. (2009a). Absolute versus relative ascertainment of pedophilia in men. *Sexual Abuse: A Journal of Research and Treatment*, 21, 431–441.
- Blanchard, R., Lykins, A. D., Wherrett, D., Kuban, M. E., Cantor, J. M., Blak, T., et al. (2009b). Pedophilia, hebephilia, and the DSM-V. *Archives of Sexual Behavior*, 38, 335–350.
- Blanchard, R., Watson, M. S., Choy, A., Dickey, R., Klassen, P., Kuban, M., et al. (1999). Pedophiles: Mental retardation, maternal age, and sexual orientation. *Archives of Sexual Behavior*, 28, 111–127.
- Blough, D. S. (1969). Generalization gradient shape and summation in steady-state tests. *Journal of the Experimental Analysis of Behavior*, 12, 91–104.
- Carlstedt, A., Nilsson, T., Hofvander, B., Brimse, A., Innala, S., & Anckarsäter, H. (2009). Does victim age differentiate between perpetrators of sexual child abuse? A study of mental health, psychosocial circumstances, and crimes. *Sexual Abuse: A Journal of Research and Treatment*, 21, 442–454.
- Chivers, M. L., Seto, M. C., & Blanchard, R. (2007). Gender and sexual orientation differences in sexual response to sexual activities versus gender of actors in sexual films. *Journal of Personality and Social Psychology*, 93, 1108–1121.
- Frenzel, R. R., & Lang, R. A. (1989). Identifying sexual preferences in intrafamilial and extrafamilial child sexual abusers. *Annals of Sex Research*, 2, 255–275.
- Freund, K. (1967). Diagnosing homo- or heterosexuality and erotic age-preference by means of a psychophysiological test. *Behaviour Research and Therapy*, 5, 209–228.
- Freund, K., Langevin, R., Cibiri, S., & Zajac, Y. (1973). Heterosexual aversion in homosexual males. *British Journal of Psychiatry*, 122, 163–169.
- Freund, K., McKnight, C. K., Langevin, R., & Cibiri, S. (1972). The female child as a surrogate object. *Archives of Sexual Behavior*, 2, 119–133.
- Freund, K., Sedlacek, F., & Knob, K. (1965). A simple transducer for mechanical plethysmography of the male genital. *Journal of the Experimental Analysis of Behavior*, 8, 169–170.
- Kinsey, A. C., Pomeroy, W. B., & Martin, C. E. (1948). *Sexual behavior in the human male*. Philadelphia: W. B. Saunders.
- Kuban, M., Barbaree, H. E., & Blanchard, R. (1999). A comparison of volume and circumference phallometry: Response magnitude and method agreement. *Archives of Sexual Behavior*, 28, 345–359.
- Laumann, E. O., Gagnon, J. H., Michael, R. T., & Michaels, S. (1994). *The social organization of sexuality: Sexual practices in the United States*. Chicago: University of Chicago Press.
- Lykins, A. D., Cantor, J. M., Kuban, M. E., Blak, T., Dickey, R., Klassen, P. E., & Blanchard, R. (2010a). Sexual arousal to female children in gynephilic men. *Sexual Abuse: A Journal of Research and Treatment*, 22, 279–289.
- Lykins, A. D., Cantor, J. M., Kuban, M. E., Blak, T., Dickey, R., Klassen, P. E., et al. (2010b). The relation between peak response magnitudes and agreement in diagnoses obtained from two different phallometric tests for pedophilia. *Sexual Abuse: A Journal of Research and Treatment*, 22, 42–57.